European Journal of Soil Science, December 2000, **51,** 717–728

Estimating variograms of soil properties by the method-of-moments and maximum likelihood

R. M. LARK
Silsoe Research Institute, Wrest Park, Silsoe, Bedford MK45 4HS, UK

Summary

Variograms of soil properties are usually obtained by estimating the variogram for distinct lag classes by the method-of-moments and fitting an appropriate model to the estimates. An alternative is to fit a model by maximum likelihood to data on the assumption that they are a realization of a multivariate Gaussian process. This paper compares the two using both simulation and real data.

The method-of-moments and maximum likelihood were used to estimate the variograms of data simulated from stationary Gaussian processes. In one example, where the simulated field was sampled at different intensities, maximum likelihood estimation was consistently more efficient than the method-of-moments, but this result was not general and the relative performance of the methods depends on the form of the variogram. Where the nugget variance was relatively small and the correlation range of the data was large the method-of-moments was at an advantage and likewise in the presence of data from a contaminating distribution. When fields were simulated with positive skew this affected the results of both the method-of-moments and maximum likelihood.

The two methods were used to estimate variograms from actual metal concentrations in topsoil in the Swiss Jura, and the variograms were used for kriging. Both estimators were susceptible to sampling problems which resulted in over- or underestimation of the variance of three of the metals by kriging. For four other metals the results for kriging using the variogram obtained by maximum likelihood were consistently closer to the theoretical expectation than the results for kriging with the variogram obtained by the method-of-moments, although the differences between the results using the two approaches were not significantly different from each other or from expectation. Soil scientists should use both procedures in their analysis and compare the results.

Introduction

Geostatistics has been widely applied in soil science since Burgess & Webster (1980) introduced it. The key idea is that the value of a variable at a location $\mathbf{x}$ can be regarded as a realization of a random function $Z(\mathbf{x})$ which is intrinsically stationary. This is a weak form of second-order stationarity and is met if two conditions hold. The first is that

$$\mathrm{E}[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})] = 0 \;\forall \mathbf{x}, \tag{1}$$

where $\mathbf{h}$ is a separation in space, the *lag*. The second is that the variance of the differences,

$$2\gamma(\mathbf{h}) = \mathrm{E}[\{Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})\}^2], \tag{2}$$

depends only on $\mathbf{h}$ and not on $\mathbf{x}$. The function $\gamma(\mathbf{h})$ is the *variogram*.

E-mail: murray.lark@bbsrc.ac.uk
Received 28 February 2000; revised version accepted 1 June 2000

The variogram is central in geostatistics. If the variogram of a soil property in a region can be estimated then it can be used to estimate that property over a block or at an unvisited site as a weighted combination of observed values, achieving a minimum error variance. This is the technique of kriging. Furthermore, when the variogram is known a grid survey can be designed for soil properties which will achieve a target error variance at minimum cost (McBratney *et al.*, 1981). The variogram can also be used for simulating spatial fields with the same statistical properties as an actual soil variable (Papritz & Webster, 1995) and to gain insight into the spatial structure of a set of data (Oliver, 1999).

Conventionally the variogram is obtained from a set of observed values in two stages, described by Webster & Oliver (2000). In the first stage the variogram is estimated for separate lags. The resulting set of estimates is sometimes called the experimental variogram. If we have $N_{\mathbf{h}}$ pairs of observations separated by lag $\mathbf{h}$, namely $\{z(\mathbf{x}_i), z(\mathbf{x}_i+\mathbf{h})\}$, $i = 1, 2, \ldots, N_{\mathbf{h}}$, then the variogram $\gamma(\mathbf{h})$ is estimated by

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N_{\mathbf{h}}} \sum_{i=1}^{N_{\mathbf{h}}} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2. \quad (3)$$

In practice our data will often come from sampling sites that are not laid out in a strictly regular way. In these circumstances there may only be one or a very few pairs of observations that are exactly separated by a particular lag. It is necessary to estimate the variogram for lag classes, each centred on a distance and direction which define the nominal lag, but actually including a range of distances and directions. Webster & Oliver (2000) describe this in more detail. It is often possible and sometimes necessary to ignore the direction component of lag, and to define the lag class on distance alone.

Choosing the lag classes well is important. If they are too narrow then the experimental variogram will be noisy, but if the classes are too broad then the experimental variogram may be excessively smoothed and information about the spatial structure of the variable may be lost. The procedure may be helped by interactive software (Pannatier, 1996), but is always somewhat *ad hoc* and so rather unsatisfactory.

The second stage is to fit a continuous function of lag to the experimental variogram. This enables semivariances to be calculated for all the lags in the kriging equations. Only certain mathematical functions are suitable for this purpose and choosing and fitting a model must be done with care. Webster & Oliver (2000) describe the most commonly used functions and how they are fitted.

The procedure outlined above, the method-of-moments estimation using Equation (3) followed by model-fitting, is widely used and has been studied in some detail by soil scientists. Webster & Oliver (1992) addressed the question of how many data are needed to obtain a reliable estimate of the variogram. They concluded that some 100–150 data are needed as a minimum to estimate an isotropic variogram and that a sample of around 225 data is ideal. Lark (2000) considered the usefulness of robust alternatives to the estimator in Equation (3) in soil science. A recent paper by Pardo-Igúzquiza & Dowd (1998) suggested that an alternative method for estimating the variogram, based on the method of maximum likelihood, might be used in soil science.

Maximum likelihood (ML) estimation of a model of the spatial covariance of a random variable was proposed by Mardia & Marshall (1984) in the context of regression analysis and by Kitanidis (1983, 1987) for specifically geostatistical purposes. To apply the basic ML procedure we have to assume that the n data, $z(\mathbf{x}_i)$, $i = 1, 2, \ldots, n$, are a realization of a stationary, n-variate Gaussian process of variance σ^2 and variance-covariance matrix $\mathbf{V}$. We assume that element (i, j) of $\mathbf{V}$ equals $C(\mathbf{x}_i - \mathbf{x}_j)$, where $C(\mathbf{h})$ is the auto-covariance function. The auto-covariance function is defined as

$$C(\mathbf{h}) = \mathrm{E}[\{Z(\mathbf{x}) - \mu\}\{Z(\mathbf{x} + \mathbf{h}) - \mu\}], \quad (4)$$

where $\mu = \mathrm{E}[Z(\mathbf{x})]$, assumed to be constant for all $\mathbf{x}$. This latter assumption, and the assumption that the auto-covariance function in Equation (4) depends on $\mathbf{h}$ only and not on $\mathbf{x}$, constitutes the condition of second-order stationarity. It is a more restrictive assumption than the intrinsic hypothesis, Equations (1) and (2), which is all that we have to assume to define the variogram and to estimate it by the method-of-moments (Webster & Oliver, 2000). Under the assumption of second-order stationarity the variogram defined in Equation (2) may be written as

$$\gamma(\mathbf{h}) = \sigma^2 - C(\mathbf{h}). \quad (5)$$

The joint probability density function of the data, assumed to be multivariate Gaussian, is written as:

$$p(\mathbf{z}, \mathbf{m}, \mathbf{p}) = (2\pi)^{-\frac{n}{2}} |\mathbf{V}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{z} - \mathbf{m})^{\mathrm{T}} \mathbf{V}^{-1} (\mathbf{z} - \mathbf{m})\}. \quad (6)$$

The vector $\mathbf{z}$ contains the n data, and $\mathbf{m}$ is a vector of length n, all the elements of which are equal to the mean of the process. The vector $\mathbf{p}$ contains parameters of the covariance matrix.

The matrix $\mathbf{V}$ can be thought of as a product of a scalar, the variance σ^2, and the auto-correlation matrix $\mathbf{A}$ which is used for convenience:

$$\mathbf{V} = \sigma^2 \mathbf{A}. \quad (7)$$

Let us assume that we wish to fit to the data a variogram function that consists of a nugget component and a spatially structured component described by $f(\mathbf{h}|r)$, a conditional negative semi-definite function of lag with a single spatial parameter, r (e.g. an exponential or spherical function):

$$\gamma(\mathbf{h}) = c_0 + c\ f(\mathbf{h}|r). \quad (8)$$

The expression $f(\mathbf{h}|r)$ either reaches 1 at some value of $\mathbf{h}$, or tends to 1 as $\mathbf{h} \to \infty$. Element (i, j) of $\mathbf{A}$ is then given by

$$\begin{aligned} \mathbf{A}(i,j) &= 1 & i = j \\ &= s\{1 - f(\mathbf{x}_i - \mathbf{x}_j|r)\} & i \neq j \end{aligned}, \quad (9)$$

where s is the spatial dependence, the proportion of the variance of the random function which has a spatial structure described by the variogram,

$$s = \frac{c}{c_0 + c}. \quad (10)$$

These two parameters, r and s, describe the matrix $\mathbf{A}$. A similar parameterization could be obtained for more complex spatial patterns such as nested random functions with components of different range.

If the probability density function in Equation (6) is regarded as a function of $\mathbf{m}$ and $\mathbf{p}$, with data $\mathbf{z}$ regarded as fixed, then it defines a likelihood function. The values of $\mathbf{m}$ and $\mathbf{p}$ that maximize the function for a given set of data are maximum likelihood estimates of the parameters, the set of

parameters such that the observed data have the largest probability of occurring. For practical purposes the negative log-likelihood function is normally used, and a minimum value is found. For the multi-Gaussian distribution with the covariance matrix factorized into a scalar variance and a matrix **A** defined in terms of a variogram function, the negative log-likelihood function may be written as

$$L(\mathbf{m}, \sigma^2, r, s|\mathbf{z}) = \frac{n}{2}\ln(2\pi) + n \ln \sigma + \frac{1}{2}\ln|\mathbf{A}| + \frac{1}{2\sigma^2}(\mathbf{z}-\mathbf{m})^{\mathrm{T}}\mathbf{A}^{-1}(\mathbf{z}-\mathbf{m}). \quad (11)$$

Maximum likelihood estimates of the vector **m** and of the variance σ^2 can be obtained by setting to zero the partial derivatives of Equation (11) with respect to these variables. The resulting estimators are

$$\hat{\mathbf{m}} = (\mathbf{1}_n^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{1}_n)^{-1}\,\mathbf{1}_n^{\mathrm{T}}\,\mathbf{A}^{-1}\,\mathbf{z}, \quad (12)$$

where $\mathbf{1}_n$ is a vector of elements all set to 1, and

$$\hat{\sigma}^2 = \frac{(\mathbf{z}-\mathbf{m})^{\mathrm{T}}\mathbf{A}^{-1}(\mathbf{z}-\mathbf{m})}{n}. \quad (13)$$

The estimates, $\hat{\mathbf{m}}$ and $\hat{\sigma}^2$ may then be substituted for **m** and σ^2 in Equation (11) to give a negative log-likelihood function of the two parameters of **A** conditional on the estimates:

$$L(r, s|\hat{\mathbf{m}}, \hat{\sigma}^2, \mathbf{z}) = \frac{n}{2}\ln(2\pi) + \frac{n}{2} - \frac{n}{2}\ln(n) + \frac{1}{2}\ln|\mathbf{A}| + \frac{n}{2}\ln\left\{(\mathbf{z}-\hat{\mathbf{m}})^{\mathrm{T}}\mathbf{A}^{-1}(\mathbf{z}-\hat{\mathbf{m}})\right\}. \quad (14)$$

Maximum likelihood estimates of r and s, and so of the underlying variogram, may then be obtained by finding the values that minimize this function.

If the assumption of a constant mean, $\mu = \mathrm{E}[Z(\mathbf{x})]$, is implausible then the ML method may be extended to estimate simultaneously the parameters of a trend model and a variogram of the residuals from the trend. This is best done by restricted maximum likelihood (REML). Pardo-Igúzquiza (1997) discusses the procedure and presents Fortran code, but I do not consider REML further in this paper.

Estimation of the variogram by ML methods is controversial. One of the main objections (see, for example, Cressie, 1993) is the assumption of multivariate normality which is explicit in the definition of the negative log-likelihood function in Equation (11) (although Kitanidis (1985) reported studies with non-normal simulated data where ML estimation of the variogram outperformed other methods). Many soil properties are not normally distributed on the scales on which they are most naturally or conveniently measured. Non-Gaussian behaviour may be identified by exploratory data analysis using graphs (e.g. histograms) and third- and fourth-order moments of the data (coefficients, respectively, of skew and kurtosis). When non-normality is found data can often be made approximately Gaussian by transformation to a new scale (e.g. by taking logarithms or square-roots), and this is widely followed as best practice in the analysis of soil data (Webster & Oliver, 1990).

Even if we have transformed a set of n data $z_1, \ldots, z_n$ to univariate normality, this is only a necessary and not a sufficient condition for the assumption that they are a realization of an n-variate Gaussian process. Pardo-Igúzquiza (1998a) noted that the actual multivariate distribution which underlies a set of data can never be verified since only one realization is available. He went on to argue that, given this uncertainty, the multivariate normal distribution is a natural assumption because it is the distribution of maximum entropy when all that is known is the mean and covariance matrix (which fully characterize a multi-Gaussian process). Furthermore, he pointed out that the negative log-likelihood function for a data set, given the estimated variance, has two terms which have a general meaning. These are

$$\ln|\mathbf{A}| \quad \text{and} \quad n \ln\{(\mathbf{z}-\mathbf{m})'\mathbf{A}^{-1}(\mathbf{z}-\mathbf{m})\}. \quad (15)$$

The first term is a general measure of spatial uncertainty with its maximum value for a pure nugget process ($\ln|\mathbf{A}| = 0$) and any other value ($\ln|\mathbf{A}| < 0$) implying some degree of spatial correlation. The second term is a weighted least-squares criterion for the fit of the model. The set of variogram parameters that minimizes Equation (11), and hence also Equation (15), could therefore be regarded as a rational choice without making any assumptions about the multivariate distribution. In selecting that set we opt for a model with strong spatial structure without incurring an excessive penalty from the squared error term. Pardo-Igúzquiza (1998a) argued that the weighting of the two terms (1 and n, respectively) is optimal by an entropy criterion.

Another problem with maximum likelihood estimation of the variogram is that the likelihood function may have undesirable properties, particularly when a spherical model is specified (Ripley, 1988; Mardia & Watkins, 1989). Warnes & Ripley (1987) showed that the negative log-likelihood function may have several local minima and a long shallow profile, both of which make difficult the identification of a global minimum by efficient algorithms.

Estimation of the variogram by ML is computationally intensive. For n data we must invert an $n \times n$ matrix for each evaluation of the likelihood function. Many evaluations of the function will be necessary when searching for a minimum, even with efficient algorithms. Since the introduction of the method it has been recognized that it is practical for data only where $n < 150$ (Mardia & Marshall, 1984; Kitanidis, 1987). An approximation to the negative log-likelihood function was proposed by Vecchia (1988) and implemented in an algorithm by Pardo-Igúzquiza & Dowd (1997). The method uses

covariance matrices for subsets of the data to build an approximate likelihood function. This reduces the number of computational steps by a factor of 1000. I am not aware of any evaluation of this method for bias or loss of efficiency arising from the approximation.

Arguments have been advanced in favour of ML estimation of the variogram. Pardo-Igúzquiza (1998a) states that it will be more efficient than other procedures and elsewhere (Pardo-Igúzquiza, 1998b) that 'a few dozen' data may suffice (compare this with Webster & Oliver (1992) who recommend at least 100 data for the method-of-moments). A particular advantage of ML estimation is that it does not require *ad hoc* definition of lag classes, so there will be no smoothing of spatial structure.

A further argument against the method-of-moments has recently been stated by Diggle & Ribeiro (1999). Although $\hat{\gamma}(k)$ defined in Equation (3) is an unbiased estimate of the variogram, they suggest that, because a set of estimates derived from the same data for a sequence of lag classes will be strongly correlated, the sequence of estimates may give a misleading impression of the form of the variogram. This effect can be seen in the results presented by Webster & Oliver (1992). It follows that the goodness of fit of a variogram model to the empirical variogram might be a poor criterion for selecting a set of model parameters, although a new method has been proposed for fitting variogram functions (Genton, 1998) which takes account of this correlation.

The discussion above suggests that we should look at ML estimation of the variogram as a serious alternative. The possibility that ML estimation is more efficient than the method-of-moments in practice could be of particular interest if it allows variograms for soil properties to be obtained with acceptable precision from fewer data than are required by the method-of-moments.

The present paper compares the method-of-moments estimation of the variogram with ML estimation empirically. First, simulated data are used to compare the two methods with data sets of different size to see if there is reason to expect that ML estimation can work with significantly fewer data than the method-of-moments, and to compare the sensitivity of the methods to departures from assumptions. Second, the estimation methods are applied to subsets of some data on soil properties, then used for kriging to evaluate the variograms obtained by the different methods.

Analysis

Simulations

The simulation study has three objectives. First, it compares the performance of maximum likelihood and the method-of-moments for estimating the variogram using data sets of different size. Second, it compares the performance of the estimators on data with different spatial structures. Third, it evaluates the effects of skew and contamination on performance of the estimators.

The basic simulation procedure was as follows. A random Gaussian field was simulated on a square grid of 300 × 300 nodes as a realization of a random function with an isotropic spherical variogram:

$$\begin{aligned}\gamma(h) &= c_0 + c\left\{\frac{3}{2}\frac{h}{a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right\} \quad \text{for} \quad 0 < h \leq a \\ &= c_0 + c \quad \text{for} \quad h \geq a \\ &= 0 \quad \text{for} \quad h = 0,\end{aligned} \tag{16}$$

where a is the correlation range of the process, from 2 to 6 grid units in this simulation. The simulation was done using the SASIM procedure in the GSLIB library (Deutsch & Journel, 1992). The field was then sampled with a set of randomly located, non-intersecting transects each of 15 adjacent nodes along a row or column of the field, as in the study of Webster & Oliver (1992). The variogram was then estimated from the data in the sample using Equation (3) for lag intervals of 1, 2, …, 7 grid units. These point estimates were then output for model-fitting.

Variograms were then fitted to the transect data from the sample by maximum likelihood. In this case we know that the model of the underlying process is the spherical, but in practice we do not. Simply to fit spherical models to these data would exclude an element of uncertainty which is present when analysing actual soil data. In practice several models would be fitted (perhaps after inspecting the experimental variogram) and the best-fitting one would be selected. Thus both the choice of model and the fitted parameters are sources of uncertainty. To introduce this into the comparison between the method-of-moments and the ML we would ideally assume ignorance about the underlying process and try a range of models. It was impractical here to compare the fit of all commonly used bounded variogram functions. For this reason two options were considered, the spherical (Equation (16)) and the exponential variogram model:

$$\gamma(h) = c_0 + c\left\{1 - \exp\left(-\frac{h}{a}\right)\right\}. \tag{17}$$

The fitting was done by numerical minimization of the conditional negative log-likelihood functions as given in Equation (14). A direct grid search was performed over the two-dimensional space defined by the spatial parameter a and the spatial dependence s, Equation (10). The matrix $\mathbf{A}$ was specified from each combination of these parameters, the mean and variance were then estimated using Equations (12) and (13), then the estimate $\hat{\mathbf{m}}$ and $\mathbf{A}$ were substituted into Equation (14) to obtain L. Having fitted both variogram functions in this way the one with the smallest minimized value of the likelihood function was selected, since the models have the same number of parameters.

The procedure above generated a set of point estimates of the variogram obtained by the method-of-moments and a model fitted by maximum likelihood from each set of transects. Another set of transects was then randomly selected, and the procedure was repeated. This was iterated until 100 sets of transects had been obtained. The sampling was done without replacement, so that no node of the grid appears in more than one of the 100 sets of transects.

Variogram functions were then fitted to each of the 100 sets of estimates $\hat{\gamma}(h)$, $h = 1, 2, \ldots, 7$ grid units. This was done automatically by weighted least squares. Again both an exponential and a spherical model were fitted and the model with the smallest residual mean square was selected. The resulting sets of estimated variogram parameters were scanned for unusual values, but there did not seem to be any problems attributable to the fitting step. A set of typical results for four sets of eight random transects (120 data) are shown in Figure 1. The maximum likelihood variograms are shown along with the method-of-moments estimates of the variogram and fitted model.

The procedure above generates a set of 100 pairs of variogram parameters for the sampled field. We need an appropriate criterion for the closeness of a given model to the specified variogram used to simulate the field; which is independent of the model type since not all the fitted models will be of the specified type.

The criterion used for this purpose is the δ–neighbourhood, proposed by Diamond & Armstrong (1984) as a measure of the proximity of two variogram functions. Let $\mathcal{G}$ be the set of valid variogram functions, and let $\gamma \in \mathcal{G}$ be the variogram of a particular spatial process. The δ–neighbourhood of γ, $\mathcal{N}_\delta(\gamma)$ is a subset of $\mathcal{G}$ defined by

$$\mathcal{N}_\delta(\gamma) = \left\{ g \in \mathcal{G} : \left| \frac{g}{\gamma} - 1 \right| < \delta \right\}, \tag{18}$$

where $|\cdot|$ denotes the maximum value of the term enclosed for continuous functions from 0 to ∞.

Diamond & Armstrong (1984) showed that if sampling or modelling error results in a variogram g being used for kriging from data that are a realization of a process with the variogram γ such that

$$g \in \mathcal{N}_\delta(\gamma), \tag{19}$$

then limits on the error of the kriged estimates of the variable and its variance which are due to error in the variogram can be written in terms of δ. The δ–neighbourhood is therefore a practically meaningful measure of the proximity of two variogram functions. The proximity of each estimated variogram to the variogram specified in the simulation was measured by finding $\delta_{\min}$, the minimum δ such that the δ–neighbourhood of the specified variogram includes the estimate.

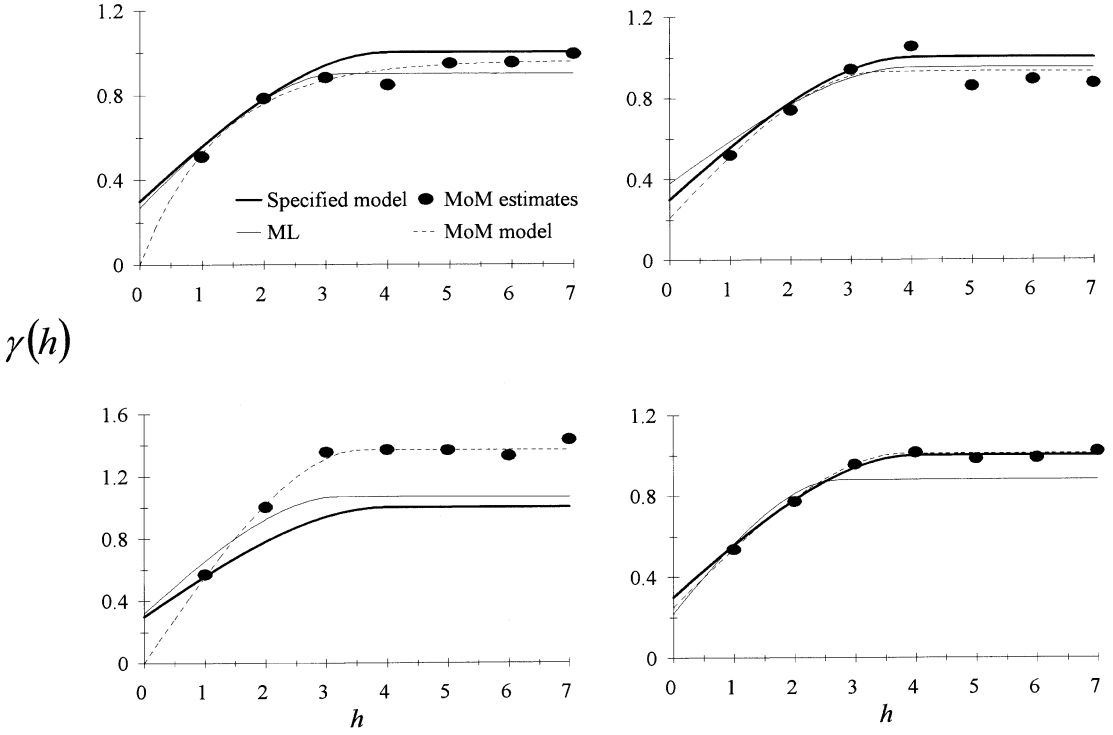

Figure 1 Four examples showing the method-of-moments estimates of the variogram from a sample of 120 data (eight transects) from a simulated random field, the model fitted automatically to these estimates and the maximum likelihood model fitted to the same data. The specified model is spherical with $c_0 = 0.3$, $c = 0.7$ and $a = 4$ grid units.

The values of δ_{min} for variograms obtained by different estimation methods from the same data will clearly depend on one another. I found for all but very small samples (one transect of 15 data) that the difference between δ_{min} for the variograms obtained by method-of-moments and maximum likelihood from the same sample was not strongly skewed. Invoking the Central Limit Theorem, the mean of this difference, over all 100 iterations of the sampling and estimation procedure, was treated as a Gaussian random variable. A paired *t*-test was carried out to test the null hypothesis that this mean difference is zero.

The basic procedure described above was applied in three experiments. In the first a Gaussian spherical process was simulated with mean zero and variogram parameters $c_0 = 0.3$, $c = 0.7$ and $a = 4$ grid units. The basic procedure was then applied to these data with sample sizes of 15 (one transect), 30 (two transects), 45, 60, 90 and 120 data.

In the second experiment the sample size was fixed at 60 (after examining the results of the first experiment) and the spatial structure of the simulated field was varied. In all cases the mean of the Gaussian random function was zero and the *a priori* variance ($c_0 + c$) was 1. All 25 combinations of the ranges {2, 3, 4, 5, 6} and nugget variances {0.1, 0.3, 0.5, 0.7, 0.9} were used.

In the third set of experiments I investigated the susceptibility of the ML procedure to non-normality of the random function and to the presence of outliers. Realizations of a random function with a non-Gaussian distribution were generated using the procedure of Lark (2000). Ninety thousand independent realizations of a standard Gaussian variable were first generated. A constant (around 4.0) was added so that the minimum value of the set was just positive, then each value was raised to a power α. This exponentiation gives the data a positive coefficient of skew. The data were then standardized to zero mean and unit variance. These new data were then used as the initial set in Deutsch & Journel's (1992) SASIM routine which rearranges them in a grid (here 300 × 300) as a realization of a random function with a specified variogram. In this instance a spherical variogram was specified with parameters $c_0 = 0.3$, $c = 0.7$ and $a = 4$ grid units. The data were then used for the basic procedure of sampling and variogram estimation, using samples of 90 data. This was carried out four times with $\alpha = 1$, 2.75, 5.0 and 7.0. The coefficients of skew for the exhaustive data were, respectively, 0.0, 1.3, 5.0 and 6.4.

In order to investigate the effects of outliers in the data, one realization of a Gaussian random function with the same spatial structure as those used in the first set of experiments was contaminated by substituting 5% of the data at randomly selected grid nodes with values drawn at random from a Gaussian random variable of mean 3.0 and variance 1.0. Variograms were then estimated and assessed following the basic procedure with 60 data per sample.

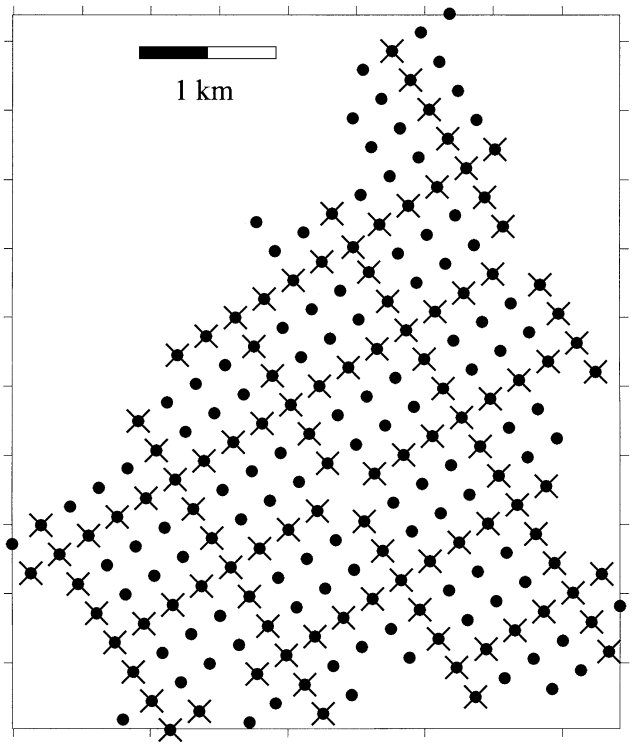

Figure 2 Location of sample sites in the Jura data. Those with crosses are the prediction set.

Studies on soil data

This part of the study aimed to compare variograms estimated from data by the method-of-moments and maximum likelihood by kriging from them. The data are heavy metal concentrations in the topsoil of a region of the Swiss Jura, measured by Atteia *et al*. (1994) and analysed fairly exhaustively by the authors (Atteia *et al*., 1994; Webster *et al*., 1994; Goovaerts *et al*., 1997). Goovaerts (1997) lists the data. Measurements were made on small cores of soil at 214 sites on a square grid of interval 250 m. The authors also measured heavy metal concentrations at clusters of points in a nested sampling design originating on points of the basic grid, but I have not used these additional points. Of the 214 data, four were excluded by Atteia *et al*. (1994) because the values were suspect. The remaining 210 data I divided into a group of 106 prediction data and one of 104 validation data. The former set consisted of 10 intersecting transects of different length such as might be used in exploratory sampling of a region. Figure 2 shows the layout of the data sets.

At each site data were available on the concentration (in mg kg^{-1}) of the soil to 25 cm depth of cadmium, cobalt, copper, nickel, lead and zinc. The data for cadmium, copper, lead and zinc were strongly positively skewed, so were transformed by taking logarithms, as discussed in the introduction. Atteia *et al*. (1994) did the same transformations of these variables. Each

metal was then considered in turn. Isotropic variograms were obtained from the prediction data. The method-of-moments was applied for lag classes centred on 250 m, 500 m, 750 m, …, 2000 m. Models were then fitted to these estimates using the MVARIOGRAM procedure in Genstat (Harding & Webster, 1995). This implements a weighted least-squares algorithm which gives greatest weight to the values of the experimental variogram for short lags. This is recommended for obtaining variograms for kriging (Webster & Oliver, 2000) and was used here because it represents standard practice. In all cases a spherical or an exponential model gave the best fit as judged by the Akaike information criterion (McBratney & Webster, 1986). Maximum likelihood estimation was also used, applying a direct search procedure to find a minimum of the negative log-likelihood function. Spherical and exponential variograms were fitted to the data, and that with the smallest minimum of the negative log-likelihood was selected.

Once a variogram was estimated it was used to predict the metal concentration at each location in the validation data set by ordinary punctual kriging from the prediction set. An estimate of the error variance for each prediction is also generated by the kriging. The predicted value at location $\mathbf{x}$ in the validation data is denoted $\hat{Z}(\mathbf{x})$ with variance $\sigma^2_{\mathrm{K},\mathbf{x}}$. This estimate can be compared with the observed value $z(\mathbf{x})$. The kriged estimate $\hat{Z}(\mathbf{x})$ is not usually very sensitive to the variogram used for kriging, but the estimated variance is sensitive. A useful statistic for validation of kriging combines both its outputs. This is $\theta(\mathbf{x})$ used by Lark (2000) and defined as

$$\theta(\mathbf{x}) = \frac{\left\{z(\mathbf{x}) - \hat{Z}(\mathbf{x})\right\}^2}{\sigma^2_{\mathrm{K},\mathbf{x}}}. \qquad (20)$$

Only the numerator of Equation (20) is a random variable, so if the kriging error is Gaussian and the variogram is correct then $\theta(\mathbf{x})$ should be distributed as χ^2 with 1 degree of freedom. Under the same conditions the kriging variance scales the expectation of $\theta(\mathbf{x})$ to 1. The value of $\theta(\mathbf{x})$ was calculated at each location in the validation data set. The mean value, $\overline{\theta}(\mathbf{x})$, should be 1 if a correct variogram has been used. However, Lark (2000) showed that the median value, $\tilde{\theta}(\mathbf{x})$, is a more useful diagnostic because although it is less powerful than the mean it is more robust to large or small data values which can have a strong effect on $\overline{\theta}(\mathbf{x})$. The median, $\tilde{\theta}(\mathbf{x})$, from a sample of $2n+1$ validation data, is distributed approximately as a Gaussian variable with an expected value of 0.455 and variance

$$\sigma^2_{\tilde{\theta}} = \frac{1}{8n\left\{\chi^2_1(0.455)\right\}^2}, \qquad (21)$$

where $\chi^2_1(0.455)$ denotes the value of the χ^2 probability density function (1 d.f.) at its median.

The median value $\tilde{\theta}(\mathbf{x})$ was evaluated for the validation data for each metal for kriging using the variograms obtained by maximum likelihood and by the method-of-moments. The two sets of values of $\theta(\mathbf{x})$ for any metal based on the two variograms constitute a set of paired samples, and were compared using a non-parametric test (Wilcoxon's signed ranks test, Siegel & Castellan, 1988). This tests the null hypothesis that the medians of the two sets of values are the same.

Results

Simulations

Figure 3 shows the mean value of $\delta_{\min}$ for the variograms based on the method-of-moments and maximum likelihood over the 100 iterations of the sampling procedure for different sample sizes. The mean δ–neighbourhood is smaller for variograms based on maximum likelihood estimation at all sample sizes, i.e. these variograms are closer to the variogram specified for the simulation, but the difference between the two variograms is significant ($P < 0.05$) only for sample sizes 60 and 90 ($P = 0.013$ in each case). Note that the mean $\delta_{\min}$ for variograms estimated by maximum likelihood from 60 data is not much larger than that for the variograms estimated by the method-of-moments from 120 data.

The mean $\delta_{\min}$ was computed from 100 iterations of a sampling scheme with 60 sample data for each of 25 simulated fields with different underlying variograms. The results are shown in Figure 4. This figure represents a space the dimensions of which are the range a of a spherical variogram and the spatial dependence s of the process (Equation (10)). Each simulated field appears as a node in this space. For more than half of these fields there was no significant difference between the mean $\delta_{\min}$ of variograms obtained by the method-of-moments and those obtained by maximum likelihood. The variograms of fields with the weakest spatial structure (short

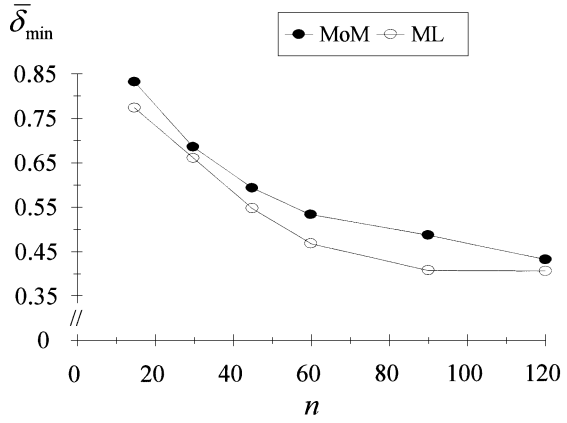

Figure 3 Mean values of $\delta_{\min}$ for variograms estimated by the method-of-moments and by maximum likelihood from samples of different size from a simulated field. The specified model is spherical with $c_0 = 0.3$, $c = 0.7$ and $a = 4$ grid units.

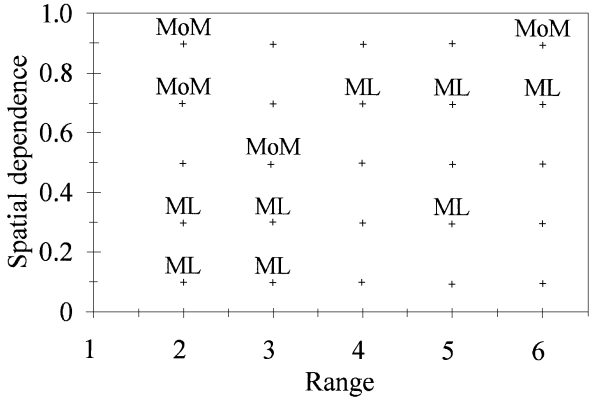

Figure 4 Spatial parameters of 25 simulated fields (spherical model, *a priori* variance of 1) with different ranges (a) and spatial dependence (Equation (10)). If a node is labelled this shows that variograms obtained from a sample of 60 data (four transects) by the indicated method had a significantly smaller mean value of $\delta_{\min}$ ($P < 0.05$) according to a paired t-test on 100 iterations of the sampling scheme.

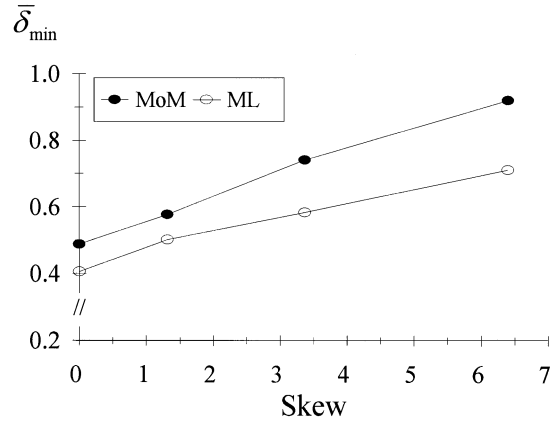

Figure 5 Mean values of $\delta_{\min}$ for variograms estimated by the method-of-moments and by maximum likelihood from samples of 90 data from simulated fields with different coefficients of skew and a common specified model, spherical with $c_0 = 0.3$, $c = 0.7$ and $a = 4$ grid units.

range and small spatial dependence) were significantly better estimated by maximum likelihood ($P < 0.05$), but some fields with longer ranges and larger spatial dependences were also best estimated this way. The method-of-moments was generally favoured on simulated data with a short spatial range and large spatial dependence.

The effects of skew on the mean value of $\delta_{\min}$ for ML or the method-of-moments are shown in Figure 5. The conformity of the exhaustive variogram of the simulated data to the specified variogram was very close in all cases, so differences in $\delta_{\min}$ can be attributed to effects of sampling and estimation. As expected the mean of $\delta_{\min}$ for the ML variograms is increased as the coefficient of skew increases and the resemblance of the data to a Gaussian variable is reduced. It is interesting that the mean of $\delta_{\min}$ for variograms estimated by the method-of-moments also increases with the data's coefficient of skew. Although the method-of-moments does not explicitly require that we assume the data to be Gaussian, using data more skew than Gaussian clearly increases its sampling error.

Contamination of the data increased the mean $\delta_{\min}$ for variograms estimated by both methods. That for the method-of-moments (60 data) was 1.41 under contamination and for maximum likelihood 1.55. This latter value is significantly larger than the former according to a paired t-test ($P = 0.019$).

Soil data

Figure 6 shows variograms obtained from the prediction data sets, and the fitted variogram models are given in Table 1. Figure 7 shows the negative log-likelihood function for two of the metals. There is a local minimum in the function for chromium at a longer range than the global minimum, and also note the long shallow 'valley bottom' in the negative log-likelihood function for cobalt. Both these features would hinder minimization by more efficient methods than the direct search.

Table 2 shows the median values, $\tilde{\theta}(\mathbf{x})$, for the metal concentrations at the validation sites obtained by kriging using the two variograms. In the case of lead and zinc $\tilde{\theta}(\mathbf{x})$ was significantly smaller than the expected value for a correct variogram (the method-of-moments variogram for zinc gives a value of $\tilde{\theta}(\mathbf{x})$ just within the 95% confidence interval). This indicates overestimation of the variance by kriging, possibly because the variograms have been affected by unusual data. In the case of nickel $\tilde{\theta}(\mathbf{x})$ is significantly larger than expected, i.e. the variance is underestimated. This could be because of the very patchy nature of the distribution of this metal which Atteia *et al.* (1994) noted. For the four remaining metals the value of $\tilde{\theta}(\mathbf{x})$ for all the variograms was not significantly different from expectation, neither was there evidence of a difference between the values of $\tilde{\theta}(\mathbf{x})$ for kriging with the two different variograms. In all cases, however, the value of $\tilde{\theta}(\mathbf{x})$ using the variogram obtained by maximum likelihood estimation was closer to the expected value of 0.455 than was the value of $\tilde{\theta}(\mathbf{x})$ obtained for kriging with the variogram estimated by the method-of-moments.

Discussion and conclusions

In the first set of simulation results the maximum likelihood estimator shows consistently better performance than does the method-of-moments, and with 60 sample data the performance of the maximum likelihood estimator is similar to that achieved by the method-of-moments with 90–120 data – often regarded as the minimum sample size required. Since estimating the variogram requires a fairly large number of data, and is therefore costly, this step can be an obstacle to the implementation of geostatistics, and so this relative efficiency

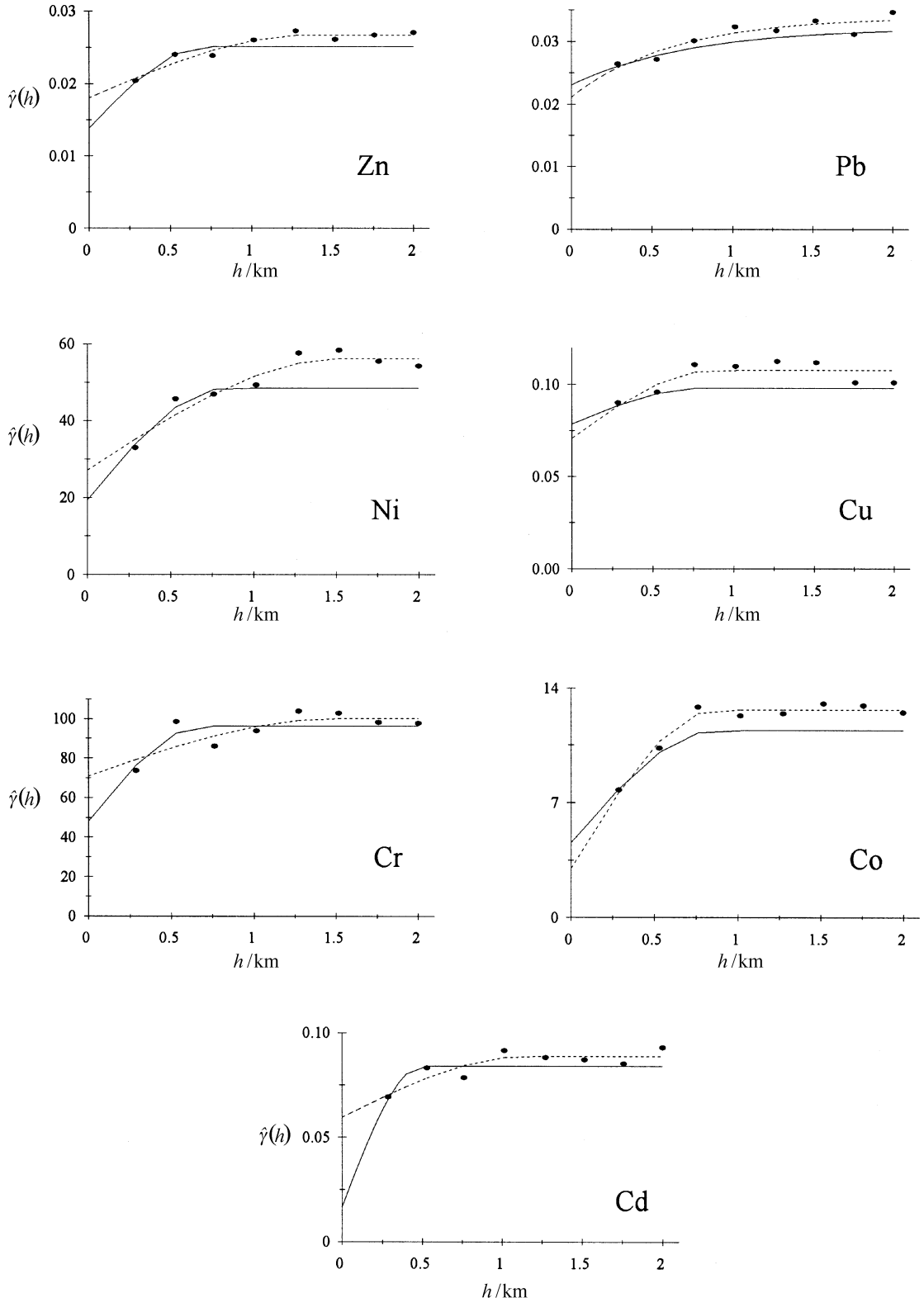

Figure 6 Variograms of soil metal concentrations fitted to the prediction data sets by the method-of-moments (broken line) or maximum likelihood (solid line).

	Method-of-moments	Maximum likelihood
Cadmium (log)	0.059 + 0.029 Sph(1.159)	0.017 + 0.067 Sph(0.503)
Cobalt	2.99 + 9.667 Sph(0.871)	4.56 + 6.84 Sph(0.868)
Chromium	70.91 + 29.24 Sph(1.497)	48.2 + 48.2 Sph (0.695)
Copper (log)	0.071 + 0.037 Sph(0.873)	0.079 + 0.020 Sph(0.798)
Nickel	27.1 + 29.2 Sph(1.542)	19.4 + 29.1 Sph(0.83)
Lead (log)	0.021 + 0.013 Exp(0.626)	0.023 + 0.009 Exp(0.73)
Zinc (log)	0.018 + 0.009 Sph(1.346)	0.014 + 0.011 Sph(0.71)

Table 1 Fitted variogram models for soil data. The models are presented in the format $c_0 + c$ model type(a). Model type is Sph (spherical) or Exp (exponential)

	$\tilde{\theta}(\mathbf{x})$ Method-of-moments	$\tilde{\theta}(\mathbf{x})$ Maximum likelihood	P
Cadmium (log)	0.319	0.402	0.98
Cobalt	0.531	0.478	0.16
Chromium	0.533	0.445	0.15
Copper (log)	0.328	0.407	0.17
Nickel	0.686	0.728	0.38
Lead (log)	0.176	0.180	0.72
Zinc (log)	0.281	0.222	0.91

The 95% confidence interval about the expected value of $\tilde{\theta}(\mathbf{x})$ (0.455) for a sample this size is 0.25–0.66.

Table 2 Values of $\tilde{\theta}(\mathbf{x})$ for kriging of metals at validation sites using different variograms. The value of P tests a null hypothesis that the distributions of $\theta(\mathbf{x})$ for the two variograms do not differ

of the maximum likelihood method is of considerable interest. Unfortunately the second set of simulation results shows that the relative performance of the two variogram estimators depends on the nature of the spatial variability. When spatial structure is weak (i.e. small spatial dependence and a short range) the maximum likelihood method seems to be advantageous, but when the spatial dependence is large the method-of-moments estimation seems to do better. It appears that the best choice of a variogram estimator depends on the form of the variogram. This accords with the findings of Zimmerman & Zimmerman (1991) who did simulation studies on processes with linear and exponential variograms.

The sensitivity of the ML method to non-normality of the data was unsurprising. It was interesting that, in the case studied (Figure 5), the method-of-moments was similarly susceptible to increased skew in the data. The advantage of ML over the method-of-moments for data with this specified variogram was not lost even when the data were substantially skew. This result is in accordance with that of Kitanidis (1985) who found that ML could outperform other estimators even on skew data. It also demonstrates the importance of using appropriate transforms on data regardless of the method used to estimate variograms.

The advantage of ML over the method-of-moments for a random function and sampling intensity where it is favoured is lost when the data are contaminated with outliers. The greater susceptibility of the maximum likelihood method to outliers is a disadvantage in practical geostatistics. Outlying values may affect geostatistical analysis of soil properties as demonstrated by Lark (2000) who showed that robust method-of-moments estimators may outperform Matheron's estimator, Equation (3), on soil data. If maximum likelihood estimation is to be used in soil science and other environmental applications then a more thorough investigation of its behaviour in the presence of outliers is necessary.

Analysing the data on the heavy metal concentrations was less conclusive. In no case was there a significant difference between the distributions of $\theta(\mathbf{x})$ for kriging with the two variograms. Both variogram estimators seemed to be susceptible to outliers leading to overestimation of the variance of kriged estimates (lead and zinc) or apparent underestimation of this variance (nickel). This reinforces the evidence from the simulation study that there is no reason to expect the maximum likelihood estimator of the variogram to be more robust in its behaviour than Matheron's method-of-moments estimator.

For the four other metals the values of $\tilde{\theta}(\mathbf{x})$ were not significantly different from expectation for either variogram estimator, but it was interesting that the result for maximum likelihood was consistently closer to the expected value of 0.455 than the result for the method-of-moments which was both larger than expected (cobalt and chromium) and smaller (copper and cadmium). Although not overwhelming evidence,

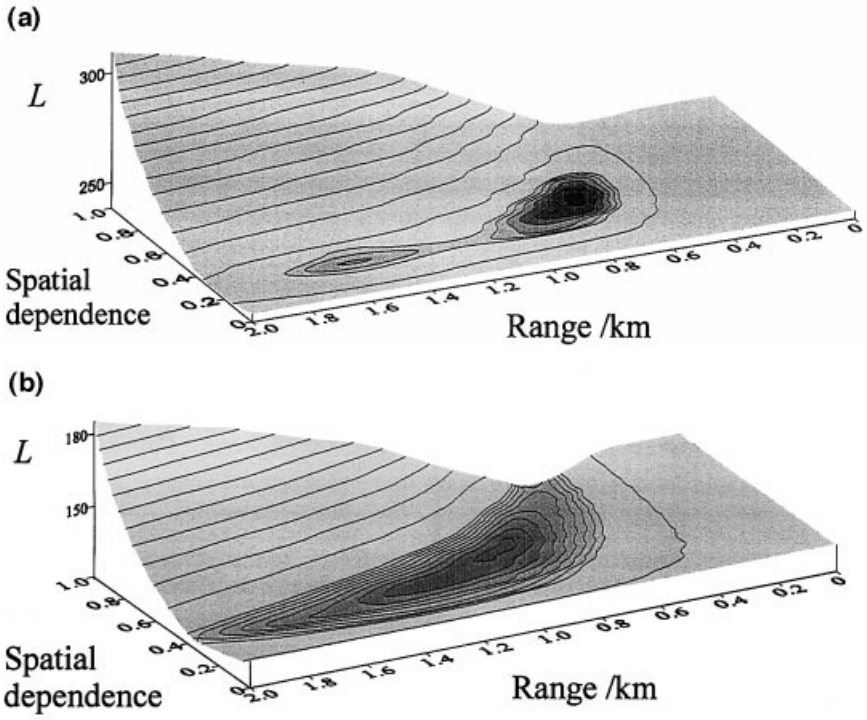

Figure 7 Negative log-likelihood functions for (a) chromium and (b) cobalt.

this does suggest that there might be advantages in using maximum likelihood to estimate variograms.

To conclude, simulation showed that in some circumstances maximum likelihood estimation of the variogram might be advantageous, but studies on soil data provided no evidence for practical advantages of maximum likelihood estimation. Nonetheless, the potential of maximum likelihood methods should be investigated further by soil scientists. I recommend that both maximum likelihood and method-of-moments estimation be used on soil data and the resulting variogram models compared. Where there are notable differences between the variograms one might be selected by cross-validation (estimating the variable at each sampled location, after excluding its datum, by kriging from the remaining data points) then comparing the kriged result to the actual data by computing $\theta(\mathbf{x})$. Lark (2000) showed that this approach can be useful for selecting among variograms obtained by several method-of-moments estimators. Cross-validation does have drawbacks, and the best procedure would be to use a separate set of validation data as in this study, but this will not be practical in routine applications as it is wasteful of data.

Other considerations might favour the use of the variogram obtained by maximum likelihood estimation. If the data from which the variograms are estimated are not laid out in a strictly regular array or set of regular transects, e.g. data from an unaligned design (Webster & Oliver, 1990), then maximum likelihood estimation may be preferred because of the smoothing effect on the variogram of using lag classes within which the actual lag distances are not tightly distributed about the mean. If the two estimators suggest that the spatial structure of the variable is weak (a nugget component comprising more than half the sill and a short range) then it is likely that the maximum likelihood estimator is at an advantage. Of course, this decision can be made only after sampling and estimation by both methods. There will be benefits from choosing the more precisely estimated variogram at this stage. Unfortunately we cannot exploit the differences in precision between the variogram estimators to achieve gains in the efficiency of sampling because until we know the form of the variogram of a soil property in a particular region we cannot identify the best estimator to use. The ML estimator does not allow us to reduce the sampling intensity to less than the recommendations made by Webster & Oliver (1992) for the method-of-moments.

A final factor should be considered. Most geostatistical analyses proceed as if the variogram – obtained by sampling, estimation and fitting of a model – is known without error. An advantage of maximum likelihood estimation is that it can be extended to obtain estimates of the variance of the variogram parameters (Pardo-Igúzquiza, 1997), which can then be used as inputs to further analysis. Dowd & Pardo-Igúzquiza (1999) showed how uncertainty about the estimated parameters of the variogram might be accounted for in simulation using their estimation variances from the ML output.

Acknowledgements

The soil data from the Swiss Jura were originally collected by Atteia *et al*. (1994) of the Swiss Federal Institute of Technology, Lausanne, and have been made available through the Internet by Dr P. Goovaerts. The work for this paper was supported by the Biotechnology and Biological Sciences Research Council of the United Kingdom through its Competitive Strategic Grant to Silsoe Research Institute.

References

Atteia, O., Dubois, J.-P. & Webster, R. 1994. Geostatistical analysis of soil contamination in the Swiss Jura. *Environmental Pollution*, **86,** 315–327.

Burgess, T.M. & Webster, R. 1980. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. *Journal of Soil Science*, **31,** 315–331.

Cressie, N.A.C. 1993. *Statistics for Spatial Data*, revised edn. John Wiley & Sons, New York.

Deutsch, C.V. & Journel, A.G. 1992. *GSLIB Geostatistical Software and User's Guide*. Oxford University Press, New York.

Diamond, P. & Armstrong, M. 1984. Robustness of variograms and conditioning of kriging matrices. *Journal of the International Association of Mathematical Geology*, **16,** 809–822.

Diggle, P.J. & Ribeiro, P.J. 1999. Bayesian inference in Gaussian model based geostatistics. In: *Proceedings of the Geostats-UK '99 Conference* (eds P.M. Atkinson, A.E. Riding & N.J. Tate), pp. 29–43. University of Southampton.

Dowd, P.A. & Pardo-Igúzquiza, E. 1999. The incorporation of model uncertainty in risk assessment using geostatistical simulation. In: *Proceedings of the Geostats-UK '99 Conference* (eds P.M. Atkinson, A.E. Riding & N.J. Tate), pp. 29–43. University of Southampton.

Genton, M.G. 1998. Variogram fitting by generalized least squares using an explicit formula for the covariance structure. *Mathematical Geology*, **30,** 323–345.

Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.

Goovaerts, P., Webster, R. & Dubois, J.-P. 1997. Assessing the risk of soil contamination in the Swiss Jura using indicator geostatistics. *Environmental and Ecological Statistics*, **4,** 31–48.

Harding, S.A. & Webster, R. 1995. Procedure MVARIOGRAM. In: *Genstat® 5 Procedure Library Manual Release 3[3]* (eds R.W. Payne & G.M. Arnold), pp. 267–269. Lawes Agricultural Trust (Rothamsted Experimental Station), Harpenden.

Kitanidis, P.K. 1983. Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Research*, **19,** 909–921.

Kitanidis, P.K. 1985. Minimum-variance unbiased quadratic estimation of covariances of regionalized variables. *Journal of the International Association of Mathematical Geology*, **17,** 195–208.

Kitanidis, P.K. 1987. Parametric estimation of covariances of regionalized variables. *Water Resources Bulletin*, **23,** 557–567.

Lark, R.M. 2000. A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of Soil Science*, **51,** 137–157.

Mardia, K.V. & Marshall, R.J. 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, **71,** 135–146.

Mardia, K.V. & Watkins, A.J. 1989. On multimodality of the likelihood in the spatial linear model. *Biometrika*, **76,** 289–295.

McBratney, A.B. & Webster, R. 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal of Soil Science*, **37,** 617–639.

McBratney, A.B., Webster, R. & Burgess, T.M. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. I. Theory and method. *Computers and Geosciences*, **7,** 331–334.

Oliver, M.A. 1999. Exploring soil spatial variation geostatistically. In: *Precision Agriculture '99, Part 1* (ed. J.V. Stafford), pp. 3–17. Sheffield Academic Press, Sheffield.

Pannatier, Y. 1996. *Variowin: Software for Spatial Data Analysis in 2D*. Springer-Verlag, New York.

Papritz, A. & Webster, R. 1995. Estimating temporal change in soil monitoring: II. Sampling from simulated fields. *European Journal of Soil Science*, **46,** 13–27.

Pardo-Igúzquiza, E. 1997. MLREML: a computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood. *Computers and Geosciences*, **23,** 153–162.

Pardo-Igúzquiza, E. 1998a. Maximum likelihood estimation of spatial covariance parameters. *Mathematical Geology*, **30,** 95–108.

Pardo-Igúzquiza, E. 1998b. Inference of spatial indicator covariance parameters by maximum likelihood using MLREML. *Computers and Geosciences*, **24,** 453–464.

Pardo-Igúzquiza, E. & Dowd, P.A. 1997. AMLE3D: a computer program for the inference of spatial covariance parameters by approximate maximum likelihood estimation. *Computers and Geosciences*, **23,** 793–805.

Pardo-Igúzquiza, E. & Dowd, P.A. 1998. Maximum likelihood inference of spatial covariance parameters of soil properties. *Soil Science*, **163,** 212–219.

Ripley, B.D. 1988. *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge.

Siegel, S. & Castellan, N.J. 1988. *Nonparametric Statistics for the Behavioural Sciences*. McGraw-Hill, New York.

Vecchia, A.V. 1988. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society B*, **50,** 297–312.

Warnes, J.J. & Ripley, B.D. 1987. Problems with likelihood estimation of covariance functions of spatial Gaussian processes. *Biometrika*, **74,** 640–642.

Webster, R. & Oliver, M.A. 1990. *Statistical Methods in Soil and Land Resource Survey*. Oxford University Press, Oxford.

Webster, R. & Oliver, M.A. 1992. Sample adequately to estimate variograms of soil properties. *Journal of Soil Science*, **43,** 177–192.

Webster, R. & Oliver, M.A. 2000. *Geostatistics for Environmental Scientists*. John Wiley & Sons, Chichester.

Webster, R., Atteia, O. & Dubois, J.-P. 1994. Coregionalization of trace metals in the soil in the Swiss Jura. *European Journal of Soil Science*, **45,** 205–218.

Zimmerman, D.L. & Zimmerman, M.B. 1991. A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics*, **33,** 77–91.